PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04J

A2

(11) International Publication Number: WO 97/04543

(43) International Publication Date: 6 February 1997 (06.02.97)

(21) International Application Number: PCT/US96/11942

(22) International Filing Date: 18 July 1996 (18.07.96)

(30) Priority Data:
60/001,498 19 July 1995 (19.07.95) US

(71) Applicant (*for all designated States except US*): ASCOM NEXION INC. [US/US]; 289 Great Road, Acton, MA 01720 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): CALDARA, Stephen, A. [US/US]; 220 Horsepond Road, Sudbury, MA 01776 (US). HAUSER, Stephen, A. [US/US]; 106 Farms Drive, Burlington, MA 01803 (US). MANNING, Thomas, A. [US/US]; 26 Summer Street, Northboro, MA 01532 (US).

(74) Agents: LEBOVICI, Victor, B. et al.; Weingarten, Schurgin, Gagnebin & Hayes, Ten Post Office Square, Boston, MA 02109 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: ALLOCATED AND DYNAMIC SWITCH FLOW CONTROL

(57) Abstract

A method and apparatus are disclosed for eliminating cell loss in a network switch through the use of flow control of both allocated and dynamic bandwidth. When output buffers in the switch become filled to a predetermined threshold level a feedback message is provided to input buffers to prevent transmission of cells from the input buffers to the output buffers. In order to provide connection and traffic type isolation the buffers are grouped into queues and flow control may be implemented on a per queue basis. The feedback message is a digital signal including an ACCEPT/REJECT message and a NO-OP/XOFF message. An XOFF message can be received while transmitting via allocated bandwith or dynamic bandwidth. In particular, an XOFF (allocated) message may be received with regard to allocated bandwidth and an XOFF (dynamic) message may be received with regard to dynamic bandwidth. An optional tagging technique may be employed to distinguish between requests for dynamic and allocated bandwidth. When ACCEPT is received by the requesting input queue the cell is transferred to the output queue. When REJECT is received by the requesting queue the cell is not transferred. When XOFF (dynamic) is received by the requesting input queue further requests to transfer to that output queue by the requesting input queue using dynamic bandwidth are halted until receipt of an XON message from that output queue. When XOFF (allocated) is received by the requesting input queue further requests to transfer to that output queue by the requesting input queue using allocated bandwidth are halted until receipt of an XON message from that output queue.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia
AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

ALLOCATED AND DYNAMIC SWITCH FLOW CONTROL

FIELD OF THE INVENTION

The present invention is generally related to telecommunications networks, and more particularly to reduction of cell loss in network switches.

RELATED APPLICATION

This application claims benefit of U.S. Provisional Application Serial No. 60/001,498, filed July 19, 1995.

BACKGROUND OF THE INVENTION

Networks such as asynchronous transfer mode ("ATM") networks are used for transfer of audio, video and other data. ATM networks deliver data by routing data units such as ATM cells from source to destination through switches. Switches include input/output ("I/O") ports through which ATM cells are received and transmitted. The appropriate output port for transmission of the cell is determined based on the cell header.

One problem associated with ATM networks is loss of cells. Cells are buffered within each switch before being routed and transmitted from the switch. More particularly, switches typically have buffers at either the inputs or outputs of the switch for temporarily storing cells prior to transmission. As network traffic increases, there is an increasing possibility that buffer space may be inadequate and data lost. If the buffer size is insufficient, cells are lost. Cell loss causes undesirable interruptions in audio and video data transmissions, and may cause more serious damage to other types of data transmissions. Avoidance of cell loss is therefore desirable.

SUMMARY OF THE INVENTION

A method and apparatus are disclosed for eliminating cell loss within a network switch through the use of flow control. The switch includes at least one input port, at least one output port, and input and output buffers associated with the respective input and output ports. Cells enter the switch through the input port and are buffered in the input buffers. The cells are then transmitted from the input buffers to the output buffers, and then transmitted to the output port. When the output buffers become filled to a predetermined threshold level, a feedback message is provided to the input buffers to prevent transmission of cells from the input buffers to the output buffers. Hence, cell loss between the input buffer and the output buffer is prevented by flow control.

In order to provide both connection and traffic type isolation, the buffers are grouped into queues and flow control is implemented on a per queue basis. Each queue includes multiple buffers, and each switch includes multiple input queues and multiple output queues. Upon entering the switch, each cell is loaded into a particular input queue for eventual transmission to a particular output queue. Individual queues are then assigned to traffic type groups in order to provide traffic type flow control if shared resources are being utilized. In an alternate implementation, each queue could be dedicated to a particular traffic type (sometimes referred to as a service class) such as the variable bit rate ("VBR") service class and the available bit rate ("ABR") service class. Flow control can then be implemented on a per traffic type basis. Further, flow control can be implemented on traffic sub-types and queues where each queue may be assigned to a particular connection, thereby providing flow control on a per connection as well as traffic subtype basis. Table 1 below shows possible flow control configurations.

TABLE 1

CONNECTION FLOW CONTROL	TRAFFIC TYPE FLOW CONTROL
No	No
Yes	No
No	Yes
Yes	Yes

Each connection is assigned bandwidth types based on the traffic type associated with the connection. There are two types of bandwidth to grant within the switch: allocated and dynamic. Allocated bandwidth is bandwidth which is "reserved" for use by the connection to which the bandwidth is allocated. Generally, a connection with allocated bandwidth is guaranteed access to the full amount of bandwidth allocated to that connection. As such, traffic types that need deterministic control of delay are assigned allocated bandwidth. Dynamic bandwidth is bandwidth which is "shared" by any of various competing connections. Because dynamic bandwidth is a shared resource, there is generally no guarantee that any particular connection will have access to a particular amount of bandwidth. For this reason dynamic bandwidth is typically assigned to connections with larger delay bounds. Other connections may be assigned a combination of dynamic and allocated bandwidth.

A digital feedback message with first and second bits is provided to facilitate flow control. The feedback message may include an ACCEPT message which can be sent from the output queue to the input queue. More particularly, using the first bit of the feedback message, first bit=0 indicates

an ACCEPT of an input queue request to transfer a cell to the output queue. When ACCEPT is received by the requesting input queue, the cell is transferred to the output queue.

The feedback message may also include a REJECT message. When REJECT is received by the requesting input queue, the cell is not transferred. However, further requests to transfer may be sent by the input queue.

The feedback message may also include a NO-OP/XOFF message. An XOFF message can be received while transmitting via allocated bandwidth or dynamic bandwidth. In particular, an XOFF (allocated) message may be received with regard to allocated bandwidth and an XOFF (dynamic) message may be received with regard to dynamic bandwidth. An optional tagging technique may be employed to distinguish between requests for dynamic and allocated bandwidth. The XOFF (dynamic) message temporarily halts transmission of requests to transfer via dynamic bandwidth. Each input queue receiving XOFF (dynamic) from a particular output queue temporarily ceases submitting requests to transmit to that particular output queue via dynamic bandwidth until a specified event occurs. The specified event could be passage of a predetermined amount of time or receipt of an XON signal which enables further requests to transfer to be sent. The input queues could also be enabled with an XON signal on a regular basis, i.e., without regard to when each particular input queue was placed in the XOFF (dynamic) state. Such a regular basis could be, for example, every 100 msec. When the second bit of the two bit message equals 0, such indicates an NO-OP (no operation) signal. Each input queue receiving a NO-OP signal is not disabled.

The XOFF (allocated) feedback message temporarily halts transmission of requests to transfer via allocated bandwidth. Each input queue receiving XOFF (allocated) from a particular output queue temporarily ceases submitting requests to transmit to that particular output queue via allocated bandwidth until a specified event occurs. The specified

event is typically receipt of an XON signal which enables further requests to transfer to be sent. The input queues could also be enabled with an XON signal on a regular basis, i.e., without regard to when each particular input queue was placed in the XOFF (allocated) state. Such a regular basis could be, for example, every 100 msec. When the second bit of the two bit message equals 0, such indicates an NO-OP signal. Each input queue receiving a NO-OP (no operation) signal is not disabled.

In the preferred embodiment an XON signal is used to enable input queues which have been placed in either XOFF state. Each input queue receiving XON from a particular output queue is enabled to submit requests to transmit to that output queue. More particularly, the XON resets both the XOFF (dynamic) and XOFF (allocated) states. The XON signal can be used in conjunction with enabling on a regular basis to both reduce unnecessary switch traffic and prevent flow blockage due to errors.

It will be apparent that various combinational responses to a request to transmit may be received by the requesting input queue. Receipt of NO-OP and either ACCEPT or REJECT operates as described above. Receipt of either XOFF (dynamic) and ACCEPT or XOFF (allocated) and ACCEPT indicates that further requests to transfer via the designated bandwidth type should cease following transfer of one cell. Receipt of XOFF (dynamic) and REJECT or XOFF (allocated) and REJECT indicates that further requests to transfer via the designated bandwidth type should cease immediately and no cells may be transmitted. Thus, the XOFF commands effect future requests while the REJECT command provides for denial of the current request.

The NO-OP/XOFF (dynamic) message is employed to reduce unnecessary feedback signaling within the switch. Switch bandwidth is inefficiently used when REJECT is repeatedly asserted when a cell can not be transmitted through the switch. XOFF (dynamic) is thus used to modify To Switch Port

Processor ("TSPP") behavior to reduce the number of requests made to a full From Switch Port Processor ("FSPP") queue.

Flow control with the feedback messages described above provides reliable point-to-multipoint transmission within the switch, i.e., transmission from a single input queue to multiple output queues. In point-to-multipoint operation the feedback messages from the multiple output queues to the single input queue are logically OR'd such that a single XOFF (dynamic) or REJECT message from any one of the plurality of output queues prevents transmission. Thus, point-to-multipoint cells are transmitted at the rate of the slowest destination queue.

Flow control with the two bit feedback messages described above also provides reliable multipoint-to-point transmission within the switch, i.e., transmission from multiple input queues to a single output queue. Each output queue has a threshold, and sends the XON message when the output queue drains to that threshold. In multipoint-to-point operation, the XON threshold of the output queue is dynamically set to reserve sufficient space for each input queue to transmit to the output queue. For example, if there are eight input queues then the threshold is set to eight so that the output queue will free sufficient space to receive all of the cells contemporaneously in serial fashion.

BRIEF DESCRIPTION OF THE DRAWING

These and other features of the present invention will be more apparent from the following detailed description in conjunction with the drawing of which:

Fig. 1 is a switch interconnect block diagram;

Fig. 2 is a block diagram illustrating point-to-point operation, switch flow control and link flow control;

Fig. 3 is a block diagram illustrating point-to-multipoint operation; and

Fig. 4 is a block diagram illustrating multipoint-to-point operation.

DETAILED DESCRIPTION OF THE DRAWING

Referring now to Fig 1, the switch includes an NxN switch fabric 10, a bandwidth arbiter 12, a plurality of To Switch Port Processor subsystems ("TSPP") 14, a plurality of To Switch Port Processor ASICs 15, a plurality of From Switch Port Processor subsystems ("FSPP") 16, a plurality of From Switch Port Processor ASICs 17 and a plurality of multipoint topology controllers 18. The NxN switch fabric, such as an ECL cross point switch fabric, is used for cell data transport, and yields N times 670 Mbps throughput. The bandwidth arbiter controls switch fabric interconnection, dynamically schedules unassigned bandwidth and resolves multipoint-to-point bandwidth contention. Each TSPP schedules transmission of cells to the switch fabric from multiple connections. Not shown are the physical line interfaces between the input link and the TSPP subsystem. The FSPP receives cells from the switch fabric and organizes those cells onto output links. Not shown are the physical line interfaces between the output link and the FSPP subsystem.

Referring now to Fig. 2, the switch includes a plurality of input ports 20, a plurality of output ports 22, and input buffers 26 and output buffers 28 associated

with the input ports and output ports, respectively. To traverse the switch, a cell 24 enters the switch through an input port and is buffered in the input buffers. The cell is then transmitted from the input buffers to output buffers in an output port. From the output port the cell is transmitted outside of the switch, for example, to another switch 29. In response to a transfer request, if the output buffers become filled to a predetermined threshold level a feedback message 30 is provided to the input ports to prevent transmission of cells from the input ports to the output buffers.

(The feedback message 30 prevents cell loss within the switch. If the number of cells 24 transmitted to the output buffers is greater than the number of available output buffers 28 then cells are lost. However, in response to a transfer request, when the output buffers 28 become filled to the threshold level the feedback message is transmitted to the input ports 20 to prevent transmission of cells from the input buffers. The threshold level is set to a value which prevents transmission of more cells than can be handled by the available output buffers. Hence, cell loss between the input buffers and the output buffers is prevented by the flow control feedback message.)

In order to provide both connection and traffic type isolation the buffers 26, 28 are organized into queues 32, 34 respectively and flow control is implemented on a per queue basis. Each queue includes multiple buffers, and each input port and output port includes multiple input queues 32 and multiple output queues 34. Upon entering the switch, each cell 24 is loaded into a particular input queue 32 for eventual transmission to a particular output queue 34. The queues are also assigned to traffic type groups in order to provide traffic type flow control if shared resources are being utilized. By assigning a unique queue per connection, flow control can then be

implemented on a per connection basis. In addition, nested queues of queues may be employed to provide per traffic type, per connection flow control.

For multipoint topologies multiple queues may be required per connection, and indirection utilized to implement per connection flow control. At the TSPP when there are multiple sources for a multipoint connection the multiple queues are nested into a scheduling list 48. At the TSPP when there is a single source a scheduling list is still employed, but having a single queue. A scheduling list is effectively a queue of queues where the queues have cells to be transmitted for that connection, and there is a scheduling list for each connection at the TSPP and the TSPP supports multiple connections. Hence, a scheduling list may be considered an input queue, and the terms are hereafter used synonymously.

Each connection is assigned bandwidth types based on the traffic type associated with the connection. There are two types of bandwidth to grant within the switch: allocated and dynamic. Allocated bandwidth is bandwidth which is "reserved" for use by the connection to which the bandwidth is allocated. Generally, a connection with allocated bandwidth is guaranteed access to the full amount of bandwidth allocated to that connection. As such, traffic types that need deterministic control of delay are assigned allocated bandwidth. Dynamic bandwidth is bandwidth which is "shared" by any of various competing connections. Because dynamic bandwidth is a shared resource, there is generally no guarantee that any particular connection will have access to a particular amount of bandwidth. For this reason dynamic bandwidth is typically assigned to connections with larger delay bounds. Other connections may be assigned a combination of dynamic and allocated bandwidth.

In order to distinguish between cells associated with connections utilizing dynamic bandwidth, allocated

bandwidth, or both, each transfer request is tagged. More particularly, transfer requests of a connection utilizing dynamic bandwidth are tagged with a bit in a first state and transfer requests of a connection utilizing allocated bandwidth are tagged with the bit in a second state. If the connection is above the allocated cell rate then the transfer request is tagged as dynamic. If the connection is operating at or below the allocated cell rate then the transfer request is tagged as allocated.

TABLE 2

Feedback Message

Bit 1	Bit 2	Meaning
0	0	ACCEPT, NO-OP
0	1	ACCEPT, XOFF (dynamic)
1	0	REJECT, NO-OP
1	1	REJECT, XOFF (dynamic)

Referring now to Fig. 1, Fig. 2 and Table 2, the feedback message 30 is provided in response to a request message 36. Prior to transmitting a cell from an input port 20 to an output port 22 the request message including the allocated/dynamic tag is sent from the input port to the output port to determine whether sufficient buffers 28 are available in the output port. The feedback message 30 provides an indication of buffer status at the output port, and transmission proceeds accordingly. The request message 36 always precedes cell transfer within the switch so that cells are only transferred under selected conditions.

In order to provide efficient flow control the feedback message 30 from the output port to the input port includes several sub-type messages. For example, the feedback message includes an ACCEPT message which may be sent in response to the request message. Using a one bit digital signal, a first bit=0 indicates an ACCEPT of an input queue request to transfer a cell to a particular output queue. When ACCEPT is received by the requesting

input queue, the cell is transferred to the output queue.

The feedback message 30 also includes a REJECT message. More particularly, the response to the request message includes either an ACCEPT or REJECT message. Using the one bit digital signal, a first bit=1 indicates a REJECT of the request to transfer a cell to the output queue. When REJECT is received by the requesting input queue, the cell is not transferred to the output queue. However, further request messages 36 may be sent from the input queue to the output queue.

In order to reduce unnecessary message traffic the feedback message 30 may also include an XOFF (dynamic) message which temporarily halts transmission of request messages 36 via dynamic bandwidth. Using a second bit of a two bit digital feedback signal, a second bit=1 indicates XOFF (dynamic). Each input queue 32 receiving XOFF (dynamic) from a particular output queue 34 temporarily ceases transmission of request messages for dynamic bandwidth to that particular output queue until a specified event occurs. The specified event could be passage of a predetermined interval of time or receipt of another signal. A second bit=0 indicates a NO-OP, i.e., a no operation message meaning that XOFF (dynamic) has not been asserted.

The feedback message may also include an XOFF (allocated) feedback message. Each input queue receiving XOFF (allocated) from a particular output queue temporarily ceases submitting requests to transmit to that particular output queue via allocated bandwidth until a specified event occurs. The specified event is typically receipt of an XON signal which enables further requests to transfer to be sent. The input queues could also be enabled with an XON signal on a regular basis, i.e., without regard to when each particular input queue was placed in the XOFF (allocated) state. Such a regular basis could be, for example, every 100 msec.

In practice the request tagging technique allows use of a single XOFF message to designate either XOFF (dynamic) or XOFF (allocated). The request tagging technique tags requests for bandwidth with a tag bit based upon whether the request is for dynamic or allocated bandwidth. The tag bit thus distinguishes allocated and dynamic requests and feedback, i.e., the XOFF transmitted in response to a request for dynamic bandwidth is XOFF (dynamic).

Utilizing the two bit feedback message, various responses to each request message may be received by the requesting input queue. Such responses are interpreted as follows. Receipt of NO-OP and either ACCEPT or REJECT operates as described above. Receipt of XOFF (dynamic) and ACCEPT indicates that further requests to transfer should cease following transfer of one cell. Receipt of XOFF (dynamic) and REJECT indicates that further requests to transfer should cease immediately and no cells may be transmitted. Thus, the XOFF (dynamic) command effects future requests while the REJECT command provides for denial of the current request.

The feedback message 30 may also include an XON message which enables further transmission of request messages. The XON message occurs asynchronously to the request to transfer messages, and is not provided in response thereto. The XON message is effective to remove both XOFF conditions. Each input queue receiving XON from a particular output queue is enabled to submit requests to transmit to that output queue.

In order to reduce the likelihood of lockup in switch flow control it may be desirable to employ timeout type functions which will allow continued operation despite the removal or failure of internal elements such as ports. For example, an input queue 32 which has ceased transmission of request messages to a particular output queue 34 following receipt of an XOFF (dynamic or

allocated) message may transmit a further request message to that output queue if an XON message is not received from that output queue within a predetermined interval of time. Alternatively, input queues may periodically transmit request messages regardless of XOFF (dynamic or allocated) state.

Referring again to Fig. 1, the invention will now be described in greater detail. In the preferred architecture each input port includes a TSPP 14, and each output port includes an FSPP 16. The TSPPs and FSPPs each include cell buffer RAM which is organized into queues 32, 34, respectively. All cells in a connection 40 pass through a single queue at each port, one at the TSPP and one at the FSPP, for the life of the connection. The queues thus preserve cell ordering. This strategy also allows quality of service ("QoS") guarantees on a per connection basis.

The request message 36 is a probe which is sent to the FSPP 16 from the TSPP 14 to determine whether sufficient buffers 34 are available for cell transmission. In order to guarantee no cell loss within the switch a TSPP cannot transmit a cell to an FSPP unless there is buffer space available for that cell. To determine buffer status, the probe communicates destination multiqueue numbers which indicate the FSPP queue or queues to which the cell is to be transmitted. For example, a destination multiqueue number could identify output queue 34a as the destination queue. When buffer space is not available in that queue, the FSPP responds to the probe with either or both of the "REJECT" and "XOFF (dynamic or allocated)" messages, as will be described below.

Three communication paths are used to implement the probe and feedback messages of switch flow control: a Probe Crossbar 42, an XOFF Crossbar 44 and an XON Crossbar 46. The Probe Crossbar 42 is an NxN crosspoint switch fabric which is used to transmit an FSPP multiqueue Number

to each FSPP. The multiqueue number identifies a plurality of destination queues for the cell for use in point-to-multipoint connections. The FSPP uses the multiqueue number to direct the probe 36 to the appropriate output queue or queues 34 and thereby determine if there are enough output buffers available in the destination queues for receipt of the cell or cells. There is a unique multiqueue number per connection per FSPP with multiple multiqueue numbers in the case of point-to-multipoint.

The XOFF Crossbar 44 is an NxN serial crosspoint switch fabric which is used to communicate "Don't Send" type messages from the FSPP 16 to the TSPP 14. Each TSPP includes multiple scheduling lists 48 which have queues of cells to be transmitted for each connection. The first bit of the feedback message 30, namely XOFF, is asserted to halt transmission of request message probes 36 from a particular TSPP's scheduling list, and is thus a state control bit which puts the receiving TSPP's scheduling list in an XOFF state, meaning that this TSPP's scheduling list 48 will not use dynamic bandwidth. This TSPP's Scheduling List then remains in the XOFF state until receiving an XON message. The second bit, namely REJECT, is asserted when insufficient buffer space is available to receive the cell in the FSPP. This situation may result from the FSPP destination queue being full or from the entire pool of output buffers being exhausted. The TSPP responds to an asserted REJECT feedback message by not dequeueing the cell 24 through the data crossbar 47. An idle cell denoted by a complemented CRC, is transmitted instead. The TSPP responds to an asserted XOFF (dynamic) feedback message by modifying the TSPP's scheduling list XOFF state bits. The XOFF state bits prevent the TSPP from attempting to send a request message from that queue on that Scheduling List until notified by the FSPPs that cell buffers are available.

The XON Crossbar 46 is an NxN serial contention-based switch which is used to communicate "Enable Send" type messages. More particularly, the XON Crossbar is employed to communicate the XON message from the FSPP to the TSPP. When the number of buffered cells in the FSPP queue drops below an XON threshold, the XON message is sent from the FSPP to the TSPP. The XON message enables the TSPP Scheduling List to resume sending request messages.

Fig. 3 illustrates point-to-multipoint switch flow control, i.e., transmission from a single input queue 32 to multiple output queues 34. In point-to-multipoint operation the XOFF crossbar performs a logical OR function. More particularly, the XOFF crossbar performs a logical OR of the feedback messages 30 asserted by the FSPPs to provided a single feedback message. As a result, receipt of REJECT or XOFF from any FSPP will cause the single resultant feedback message to be interpreted as asserting REJECT and/or XOFF respectively. This technique limits the TSPP to transmission at the rate of the slowest destination queue. However, the technique also provides desirable contemporaneous serial transmission of cells.

In the case of point-to-multipoint transmission it will be noted that a TSPP may receive multiple XON messages. Such is true because multiple XOFFs could be set by the FSPPs, i.e., more than one FSPP can assert XOFF on a transfer request. In such a case, XON messages received when the TSPP scheduling list XOFF state is clear are ignored. For example, when multiple XON messages are sent, the TSPP ignores the XONs received after the first received XON message. In the case of multipoint-to-point transmission the XON message is sent simultaneously to all TSPPs with scheduling lists transferring to an FSPP queue.

Fig. 4 illustrates multipoint-to-point switch flow control, i.e., transmission from multiple input queues 32 to a single output queue 34. Each output queue has a threshold, and the XON message is sent when the output

queue drains below that threshold. In multipoint-to-point operation, the XON threshold of the output queue is dynamically set to reserve enough buffers for each input queue to transmit to the output queue. For example, if eight queues are transmitting, the threshold is set to eight so that the output queue will free sufficient buffers to receive all eight of the cells contemporaneously in serial fashion, and thereby insuring that each queue has an opportunity to transmit.

Referring now to Figs. 1 and 4, in the case of multipoint-to-point connections, the XON crossbar 46 is used to broadcast to all TSPPs in the switch, regardless of whether or not any of the TSPPs were transmitting to the asserting FSPP queue. For the broadcast, the multipoint topology controller 18 transmits a reverse broadcast channel number on behalf of the FSPP. The receiving multipoint topology controller then performs a reverse broadcast channel to scheduling list number lookup to determine which scheduling list 48 to enable. Any TSPPs without queues transmitting to that particular FSPP queue are unaffected by the broadcast XON message since the reverse broadcast channel number look-up entry will be marked invalid.

Referring again to Fig. 2, an additional flow control enhancement provides for the queues to be organized on an hierarchical basis with multiple individual flows 52 at each hierarchical level and the feedback message 30 from the output queues to the input queues is made on the basis of the combined flow at each of the hierarchical levels. Still another enhancement provides for the queues to be organized on an hierarchical basis with multiple individual flows at each of the hierarchical levels and the feedback message from the output queues to the input queues is made on the basis of each of the individual flows.

The Probe & XOFF communication paths operate in a

pipeline fashion. First, the TSPP 14 selects an input queue 34, and information associated with that queue is used to determine output ports for transmission, i.e., a destination output queue. The bandwidth arbitrator reduces this information to a TSPP to FSPP connectivity map which is employed to control the Probe, XOFF, and data cross-points in sequence. More particularly, the FSPP multiqueue number is transmitted to the FSPP using the Probe crossbar 42. The FSPP then tests for buffer availability, and asserts REJECT and/or XOFF on the XOFF crossbar 44 if sufficient buffers are not available. The TSPP then transmits an idle cell if REJECT was asserted. If XOFF was asserted, the TSPP puts the Scheduling List into the XOFF state. If sufficient buffers are available, the TSPP transmits the cell to the FSPP output queue through the data crossbar 47.

Table 3

0=not asserted
1=asserted

Traffic Type Cell Count ≥ Limit	Queue's Dynamic Buffer Count ≥ Limit	Assert XOFF	Assert REJECT
No	No	0	0
No	Yes	1	1
Yes	No	0	1
Yes	Yes	1	1

Table 3 summarizes the policies used by the FSPP to assert REJECT and XOFF in response to requests tagged as utilizing dynamic bandwidth. The policies are based upon two relationships: the traffic type cell count in relation to the cell count limit and the queue's dynamic buffer count in relation to the buffer count limit. The traffic type cell count is a count of all cells shared by connections within a traffic type, e.g., "VBR." When the limits are not exceeded, neither REJECT nor XOFF (dynamic) is asserted. More particularly, when the traffic type cell count not greater than or equal to the limit, and the queue's dynamic buffer count is not greater than or equal to the limit, neither REJECT nor XOFF (dynamic) is asserted. When the traffic type cell count is not greater than or equal to the limit but the queue's dynamic buffer count is greater than or equal to the limit, both REJECT and XOFF (dynamic) are asserted. When the traffic type cell count is greater than or equal to the limit and the queue's dynamic buffer count is not greater than or equal to buffer limit XOFF is not

asserted but REJECT is asserted. When the traffic type cell count is greater than or equal to the limit and the queue's dynamic buffer count greater than or equal to the limit both REJECT and XOFF are asserted.

TABLE 4

1=asserted
0=not asserted

Queue's Allocated Buffer-State Count ≥ Limit	Traffic Type Cell Count ≥ Limit	Queue's Allocated Buffer Count ≥ Limit	Assert XOFF	Assert REJECT
No	No	No	0	0
No	No	Yes	1	1
No	Yes	No	0	1
No	Yes	Yes	1	1
Yes	No	No	0	1
Yes	No	Yes	1	1
Yes	Yes	No	0	1
Yes	Yes	Yes	1	1

Table 4 summarizes the policies used by the FSPP to assert REJECT and XOFF in response to requests tagged to use allocated bandwidth. The policies are based upon three relationships: the queue's allocated buffer-state count in relation to the buffer-state count limit, the traffic type cell count in relation to the cell count limit, and the queue's allocated buffer count in relation to the buffer count limit. Link flow control uses a buffer-state counter to indicate cells in flight. The allocated buffer-state

counter is used to track cells in flight for the allocated component of a connection using link flow control. Neither XOFF nor REJECT are asserted if the queue's allocated buffer-state count is greater than or equal to the count limit, the traffic type cell count is greater than or equal to the cell count limit, and the queue's allocated buffer count is greater than or equal to the count limit. Both XOFF and REJECT are asserted if the queue's allocated buffer count is greater than or equal to the buffer count limit. If the queue's allocated buffer count is not greater than or equal to the count limit, but either the queue's allocated buffer-state count is greater than or equal to the count limit or the traffic type cell count is greater than or equal to the cell count limit then REJECT is asserted and XOFF is not asserted. In all cases, if an FSPP queue has already sent an XOFF, that queue will reassert XOFF on the next probe.

It will be understood that various changes and modifications to the above described method and apparatus may be made without departing from the inventive concepts disclosed herein. Accordingly, the present invention is not to be viewed as limited to the embodiments described herein.

CLAIMS

What is claimed:

1. A method for controlling flow of at least one inputted cell within a network, the method comprising the steps of:

receiving the at least one inputted cell in a switch having a plurality of input buffers associated with a plurality of input ports, said input buffers making up at least one input queue, and an a plurality of output buffers associated with a plurality of output ports, said output buffers making up at least one output queue, the cell being received in the input queue of the input memory;

forwarding a request to transmit the at least one inputted cell to one of said at least one output queue;

granting said request to transmit if sufficient buffers are available in the output queue for receipt of the at least one inputted cell;

denying said request to transmit if sufficient buffers are not available in the output queue for receipt of the at least one inputted cell;

transmitting the at least one inputted cell from the input queue to the output queue if the submitted request is granted; and

delaying transmission of the at least one inputted cell from the input queue to the output queue if the submitted request is denied.

2. The method of claim 1 wherein said receiving step includes the further step of receiving the at least one inputted cell in one of a plurality of input queues and said transmitting step includes the further step of transmitting to at least one of a plurality of output queues.

3. The method of claim 2 wherein said submitting a request step includes the further step of submitting a request from each respective input queue in receipt of the at least one inputted cell to the respective output queues for which said respective inputted cells are destined, each request to transmit being specific to the respective input and output queues on a per transmission basis.

4. The method of claim 3 including the further step of assigning each input queue and output queue to a selected connection, whereby flow control is executed on a per connection basis.

5. The method of claim 3 including the further step of assigning each input queue and output queue to a selected service class.

6. The method of claim 3 including the further step of assigning each input queue and output queue to a selected connection and a selected service class, whereby flow control is executed on a per connection, per service class basis.

7. The method of claim 3 including the further step of providing a feedback message in response to the request to transmit, the feedback message indicating whether the submitted request is granted or denied.

8. The method of claim 7 including the further step of providing an ACCEPT/REJECT message as part of the feedback message, the at least one inputted cell in the at least one input queue being transmitted to the at least one output queue if the message indicates "ACCEPT," and not being transmitted to the at least one output queue if the message indicates "REJECT."

9. The method of claim 7 including the further step of providing an NO-OP/XOFF message as part of the feedback message, wherein each input queue receiving an XOFF from one of the plurality of output queues ceases submitting requests to transmit to the one of the plurality of output queues.

10. The method of claim 7 including, in the case of transmission from a single input queue to a plurality of output queues, the further step of performing a logical OR operation on respective ACCEPT/REJECT and NO-OP/XOFF parts of the feedback messages from the plurality of output queues such that the transmissions from the single input queue to the plurality of output queues are contemporaneous.

11. The method of claim 9 including, in the case of transmission from a plurality of input queues to a particular output queue, the further step of dynamically increasing the threshold at which the XOFF feedback message is provided by the particular output queue in order to provide sufficient buffers for contemporaneous receipt the transmissions from the plurality of input queues.

12. The method of claim 9 including the further step of providing an XON message from the one of the plurality of output queues to the one of the plurality of input queues

to re-enable the one of the plurality of input queues to submit requests to transmit to the one of the plurality of output queues.

13. The method of claim 12 including the further step of providing the XON message upon dequeueing a cell from the one of the plurality of output queues.

14. The method of claim 13 including the further step of setting a threshold in the one of the plurality of output queues for providing the XON message, the XON message being provided when the one of the plurality of output queues drains below that threshold.

15. The method of claim 12 including the further step of providing the XON message to the one of the plurality of input queues after passage of a predetermined period of time following receipt of the XOFF message.

16. The method of claim 12 including the further step of providing the XON message to the one of the plurality of input queues on a regular time basis.

17. The method of claim 12 including the further step of providing the XON message from the one of the plurality of output queues to a plurality of input queues.

18. The method of claim 2 wherein the queues include buffers, and including the further step of organizing the buffers into a hierarchy of levels with multiple individual flows of cells at each hierarchical level, wherein feedback is provided for each level based on combined flow at the respective level.

19. The method of claim 2 wherein the queues include buffers, and including the further step of organizing the

buffers into a hierarchy of levels with multiple individual flows of cells at each hierarchical level, wherein feedback is provided for each individual flow.

20. A switch for controlling flow of at least one data unit within a telecommunications network, comprising:

at least one input port for receiving inputted data units from the telecommunications network;

at least one output port for transmitting data units from said switch;

at least one input buffer queue associated with said at least one input port for temporarily storing inputted data units received at the respective input port;

at least one output buffer queue associated with said at least one input buffer queue for temporarily storing inputted data units received from said at least one input buffer queue; and

at least one feedback message generator operative to provide a feedback message to said at least one input buffer queue, said feedback message having first and second states wherein said first state indicates a grant of permission to transmit at least one data unit from one of said at least one input buffer queue to a particular output buffer queue and said second state indicates a denial of permission to transmit at least one data unit from said one of said at least one input buffer queue to said one of said at least one output buffer queue.

21. The switch of claim 14 further including at least one transmission request generator operative to request permission to transmit at least one data unit from said one of said at least one input buffer queue to said one of said at least one output buffer queue, said at least one feedback message generator providing said feedback messages in response to requests from said at least one

transmission request generator.

22. The switch of claim 15 including a feedback message generator for each input port.

23. The switch of claim 16 wherein each connection utilizes a single input queue and a single output queue, whereby flow control is executed on a per connection basis.

24. The switch of claim 16 wherein each input queue and each output queue are assigned to a selected service class, whereby flow control is executed on a per service class basis.

25. The switch of claim 16 wherein each input queue and each output queue are assigned to a selected connection and a selected service class, whereby flow control is executed on a per connection, per service class basis.

26. The switch of claim 16 wherein a feedback message is provided in response to said request to transmit, said feedback message including an indication of whether the submitted request is granted or denied.

27. The switch of claim 20 wherein said feedback message includes an ACCEPT/REJECT message, said at least one inputted cell in said at least one input queue being transmitted to said at least one output queue if the message indicates "ACCEPT," and not being transmitted to said at least one output queue if said message indicates "REJECT."

28. The switch of claim 20 wherein said feedback message includes an NO-OP/XOFF message as part of the feedback message, and wherein each input queue receiving XOFF from

one of said plurality of output queues ceases submitting requests to transmit to said one of said plurality of output queues.

29. The switch of claim 20 including, in the case of transmission from a single input queue to a plurality of output queues, a circuit which performs a logical OR operation on the respective ACCEPT/REJECT and NO-OP/XOFF parts of the feedback messages from said plurality of output queues such that transmissions from the single input queue to the plurality of output queues are serially contemporaneous.

30. The switch of claim 22 wherein, in the case of transmission from a plurality of input queues to a particular output queue, a dynamically adjustable threshold at which the XOFF feedback message is provided by the particular output queue is increased in order to provide sufficient buffers for contemporaneous receipt the transmissions from the plurality of input queues.

31. The switch of claim 22 wherein an XON message is provided from said one of the plurality of output queues to said one of the plurality of input queues to re-enable said one of the plurality of input queues to submit requests to transmit to said one of the plurality of output queues.

32. The switch of claim 25 wherein said XON message is provided upon dequeueing a cell from said one of the plurality of output queues.

33. The switch of claim 26 wherein a threshold is set in said one of the plurality of output queues for providing said XON message, said XON message being provided when the one of the plurality of output queues drains below said

threshold.

34. The switch of claim 25 wherein said XON message is provided to said one of the plurality of input queues after passage of a predetermined period of time following receipt of said XOFF message.

35. The switch of claim 25 wherein said XON message is provided to said one of the plurality of input queues on a regular time basis.

36. The switch of claim 25 wherein said XON message is provided from said one of the plurality of output queues to a plurality of input queues.

37. The switch of claim 15 wherein said queues include buffers, and said buffers are arranged in a hierarchy of levels with multiple individual flows of cells at each hierarchical level, said feedback being provided for each level based on combined flow at the respective level.

38. The switch of claim 15 wherein said queues include buffers, and said buffers are arranged into a hierarchy of levels with multiple individual flows of cells at each hierarchical level, said feedback being provided for each individual flow.

39. A method for controlling flow of at least one inputted cell within a network, the method comprising the steps of:

receiving the at least one inputted cell in a switch having a plurality of input buffers associated with at least one input queue and a plurality of output buffers associated with at least one output queue, the cell being received in the input queue of the input memory;

forwarding a request to transmit the at least one

inputted cell to one of said at least one output queue;

granting said request to transmit if sufficient buffers are available in the output queue for receipt of the at least one inputted cell;

denying said request to transmit if sufficient buffers are not available in the output queue for receipt of the at least one inputted cell;

transmitting the at least one inputted cell from the input queue to the output queue if the submitted request is granted; and

delaying transmission of the at least one inputted cell from the input queue to the output queue if the submitted request is denied.

40. The method of claim 39 including the step of, when a traffic type cell count is not greater than or equal to a first limit and the queue's dynamic buffer count is not greater than or equal to a second limit, asserting neither a REJECT nor an XOFF (dynamic).

41. The method of claim 40 including the further step of, when the traffic type cell count not greater than or equal to the first limit and the queue's dynamic buffer count is greater than or equal to the second limit, asserting both REJECT and XOFF (dynamic).

42. The method of claim 41 including the further step of, when the traffic type cell count is greater than or equal to the first limit and the queue's dynamic buffer count is not greater than or equal to the second limit, not asserting XOFF (dynamic) and asserting REJECT.

43. The method of claim 42 including the further step of, when the traffic type cell count is greater than or equal to the first limit and the queue's dynamic buffer count greater than or equal to the second limit, asserting both REJECT and XOFF (dynamic).

44. The method of claim 43 including the further step of, asserting neither an XOFF (allocated) nor a REJECT when the queue's allocated buffer-state count is greater than or equal to the count limit, the traffic type cell count is greater than or equal to the cell count limit, and the queue's allocated buffer count is greater than or equal to the count limit.

45. The method of claim 44 including the further step of asserting both XOFF (allocated) and REJECT when the queue's allocated buffer count is greater than or equal to the buffer count limit.

46. The method of claim 45 including the further step of, when the queue's allocated buffer count is not greater than or equal to the count limit, but either the queue's allocated buffer-state count is greater than or equal to the count limit or the traffic type cell count is greater than or equal to the cell count limit, asserting REJECT and not asserting XOFF (allocated).

47. The method of claim 46 including the further step of, when an FSPP queue has already sent an XOFF, reasserting XOFF on the next probe.

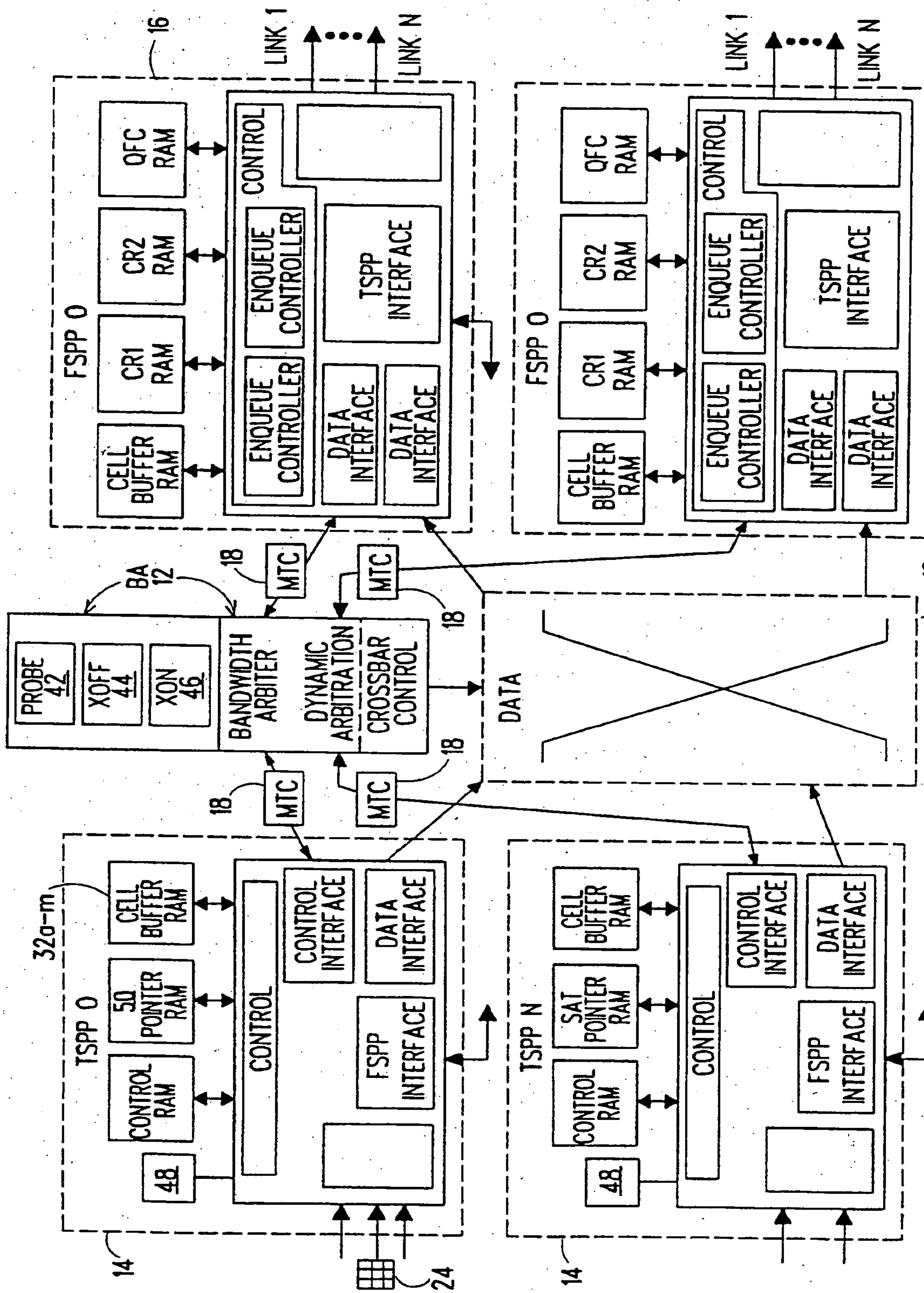
16
LINK 1
LINK N
FSPP 0
CELL BUFFER RAM
CR1 RAM
CR2 RAM
QFC RAM
ENQUEUE CONTROLLER
ENQUEUE CONTROLLER
CONTROL
DATA INTERFACE
DATA INTERFACE
TSPP INTERFACE
BA 12
18
MTC
MTC
18
PROBE 42
XOFF 44
XON 46
BANDWIDTH ARBITER
DYNAMIC ARBITRATION
CROSSBAR CONTROL
DATA
10
18
MTC
MTC
18
32a-m
TSPP 0
CONTROL RAM
50 POINTER RAM
CELL BUFFER RAM
48
CONTROL
CONTROL INTERFACE
DATA INTERFACE
FSPP INTERFACE
14
24
TSPP N
CONTROL RAM
SAT POINTER RAM
CELL BUFFER RAM
48
CONTROL
CONTROL INTERFACE
DATA INTERFACE
FSPP INTERFACE
14

FIG. 1

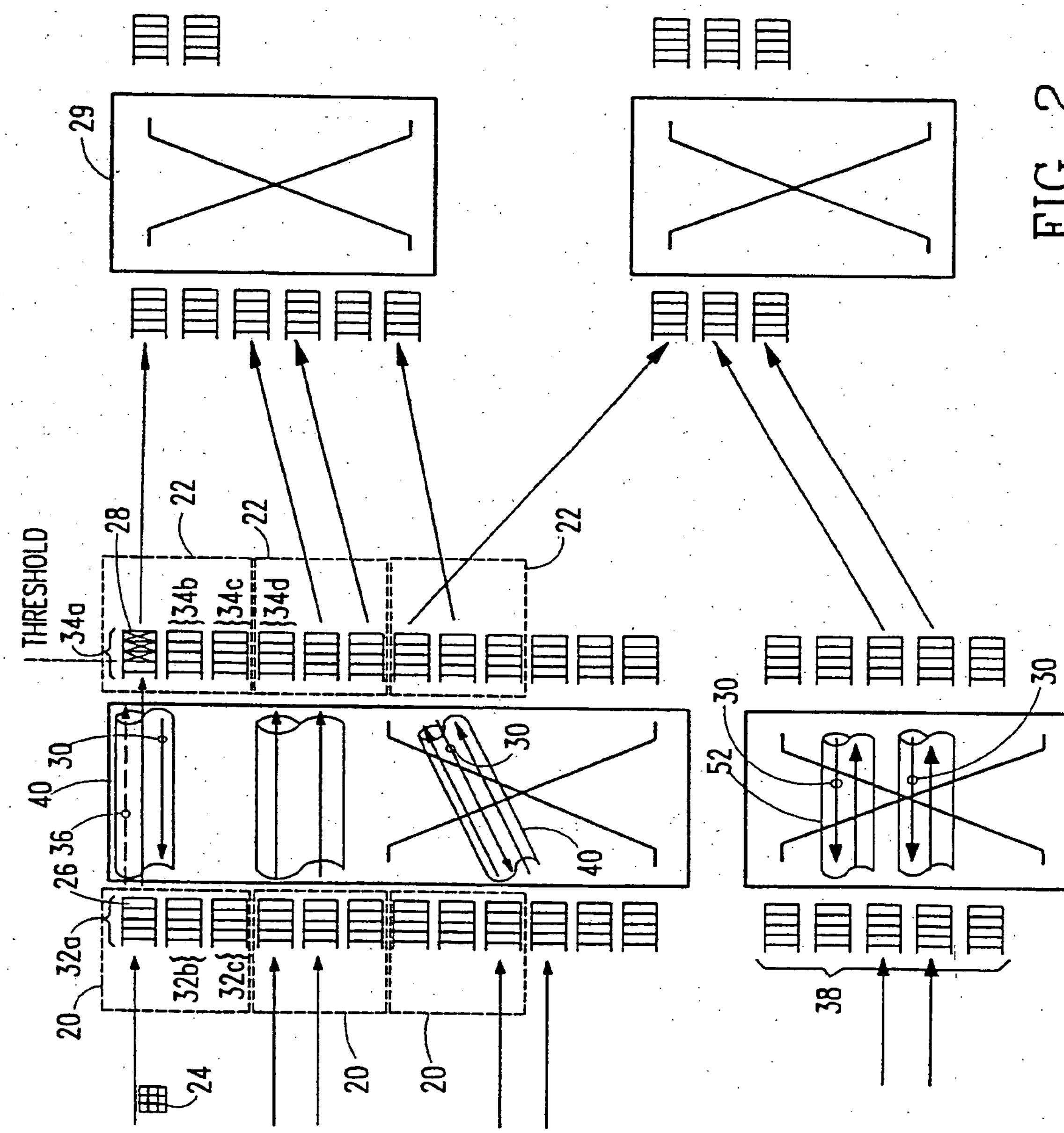
THRESHOLD
34a
28
22
22
22
34b
34c
34d
29
30
40
36
26
30
40
52
30
30
32a
32b
32c
20
20
20
24
38

FIG. 2

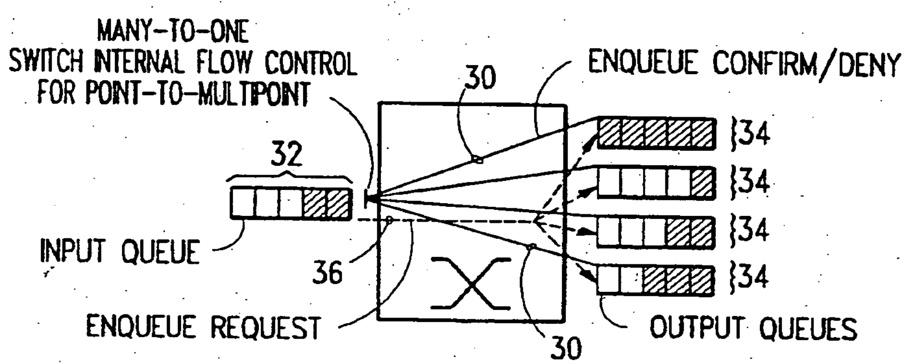
MANY-TO-ONE
SWITCH INTERNAL FLOW CONTROL
FOR POINT-TO-MULTIPOINT
30
ENQUEUE CONFIRM/DENY
32
34
34
34
34
INPUT QUEUE
36
ENQUEUE REQUEST
30
OUTPUT QUEUES

FIG. 3

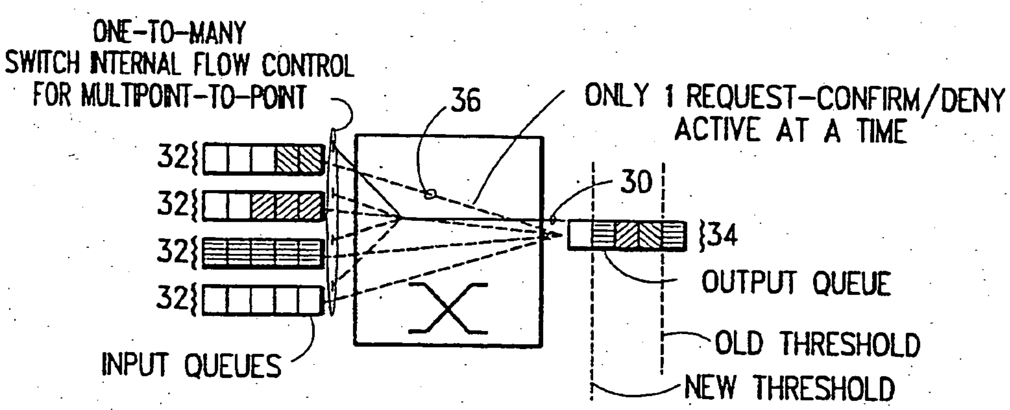
ONE-TO-MANY
SWITCH INTERNAL FLOW CONTROL
FOR MULTIPOINT-TO-POINT
36
ONLY 1 REQUEST-CONFIRM/DENY
ACTIVE AT A TIME
32
32
32
32
30
34
OUTPUT QUEUE
OLD THRESHOLD
NEW THRESHOLD
INPUT QUEUES

FIG. 4